

Homology Modeling and GPCR Drug Discovery with BioHPC

Peng Lian

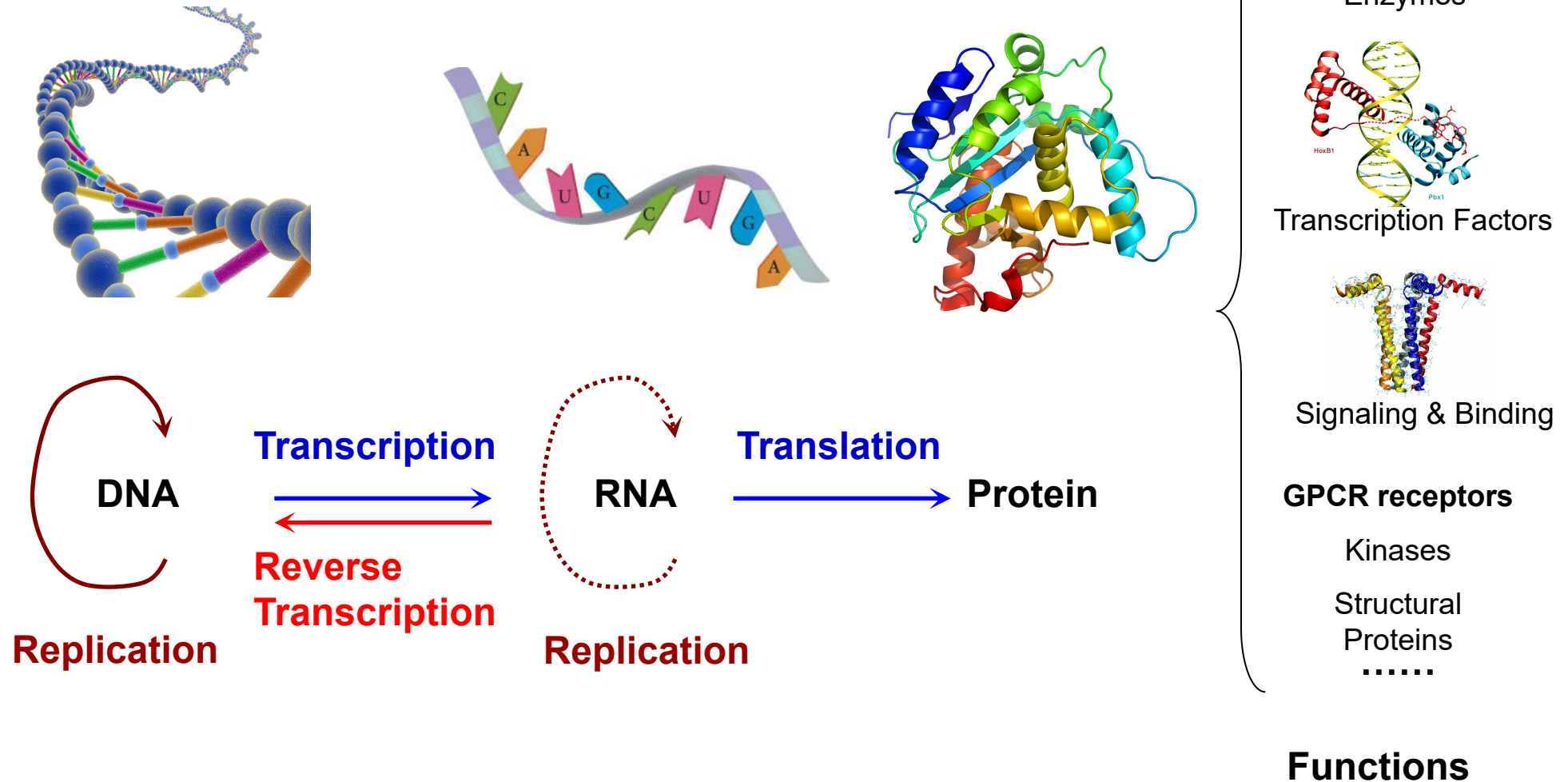
Computer Scientist, BioHPC

June 23, 2021

Outline

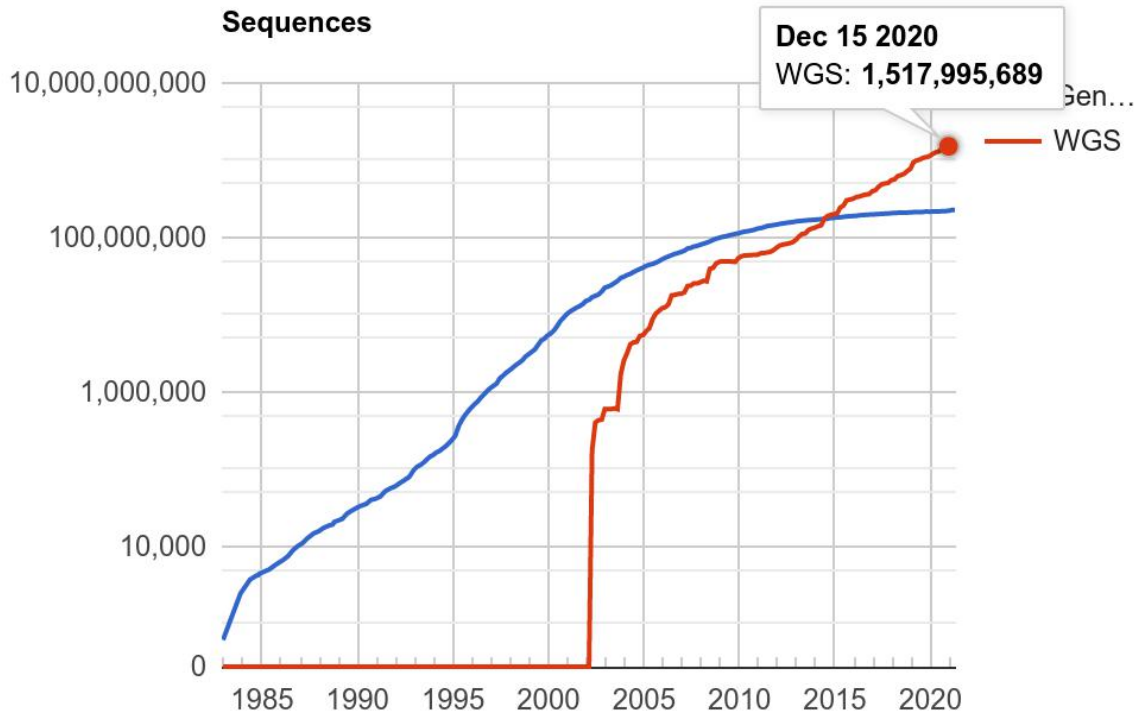
- Background on Biology, sequences, and structures
- Homology modeling
- Molecular docking
- GPCR as drug targets
- Hands-on Homology modeling of 5-HT_{1A} receptor
- Hands-on Molecular docking with Autodock on BioHPC
- Discussion on Virtual Screening

The Central Dogma of Biology



From Sequence to Structure

Sequence

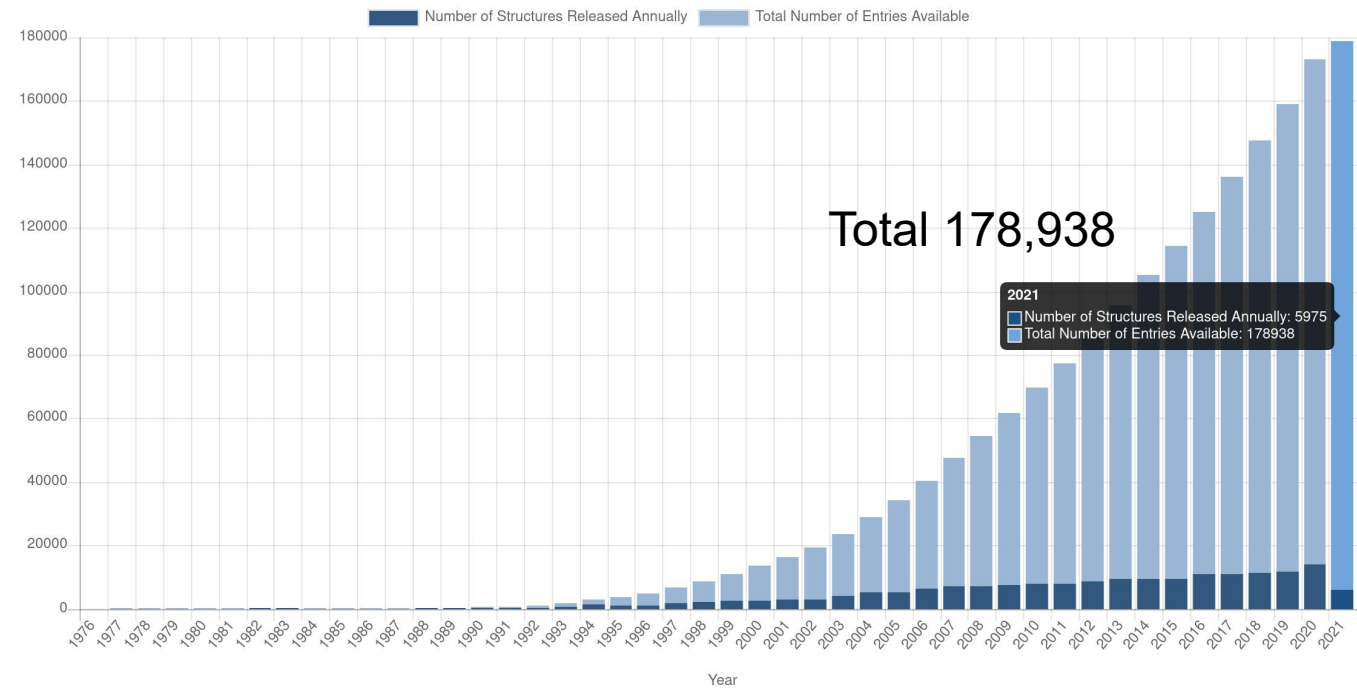


GenBank

<https://www.ncbi.nlm.nih.gov/genbank/statistics/>

Structure

DB Statistics: Overall Growth of Released Structures Per Year

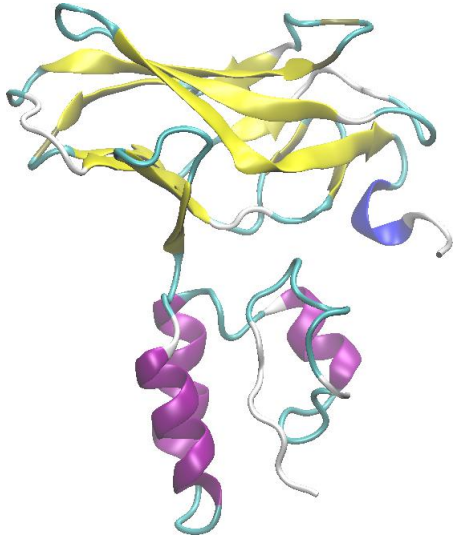


Protein Data Bank

<https://www.rcsb.org/stats/growth/growth-released-structures>

Homology Modeling

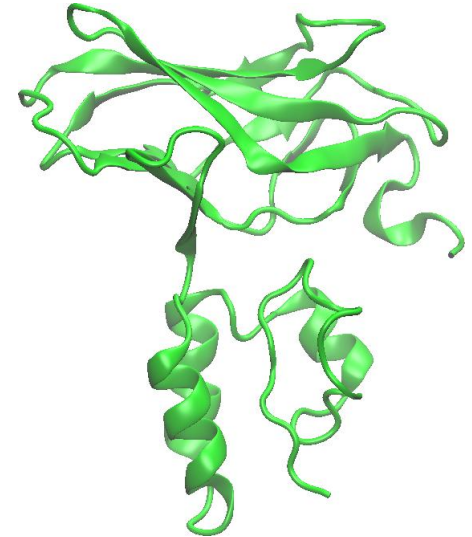
Template



Target



New Structure

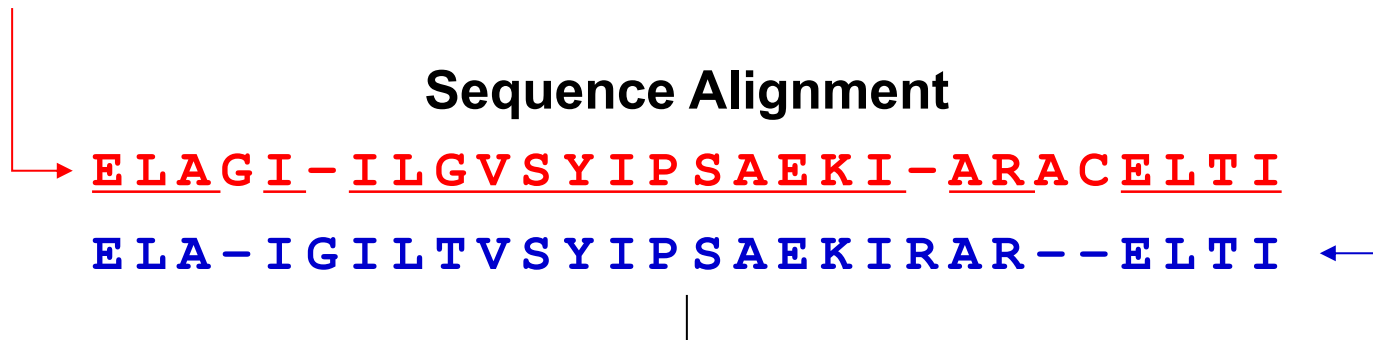


ELAGIILTVSYIPSAEKIA

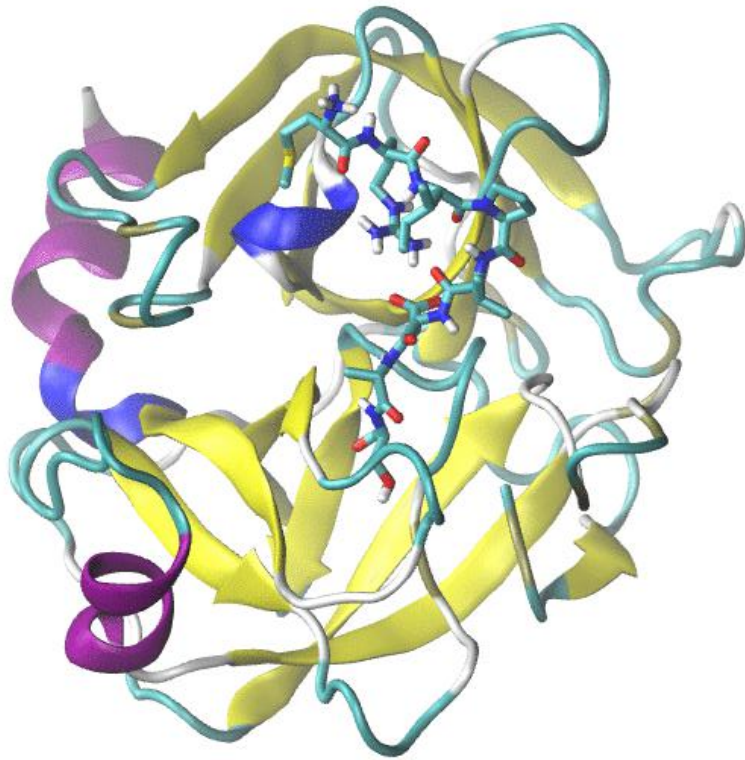
ELAIGILTVSYIPSAEKIR

ELAIGILTVSYIPSAEKIR

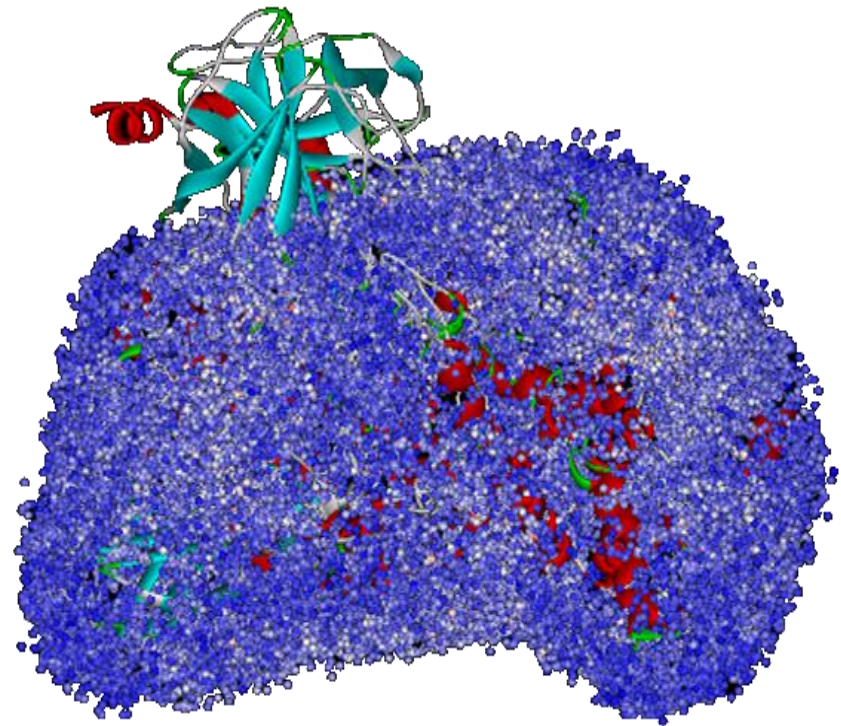
Sequence Alignment



Molecular Docking

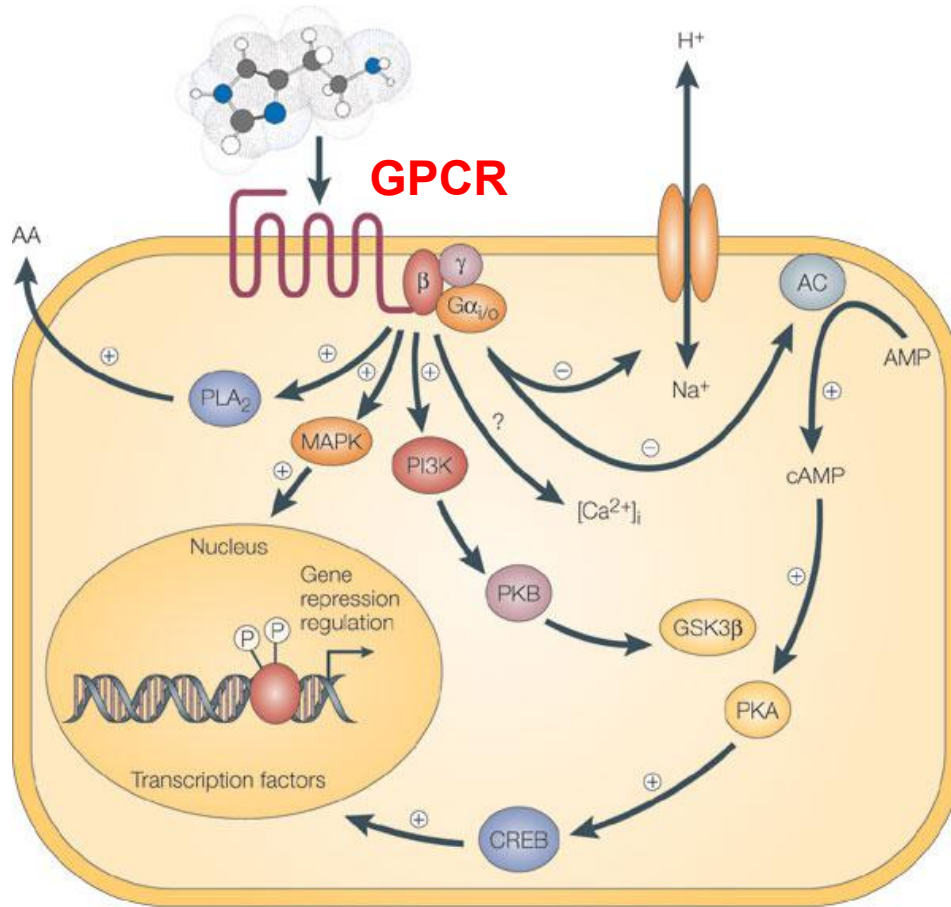


Small molecule Docking



Protein-Protein Docking

GPCR and Signaling Cascades



GPCR

Ligands

Active state

Agonists

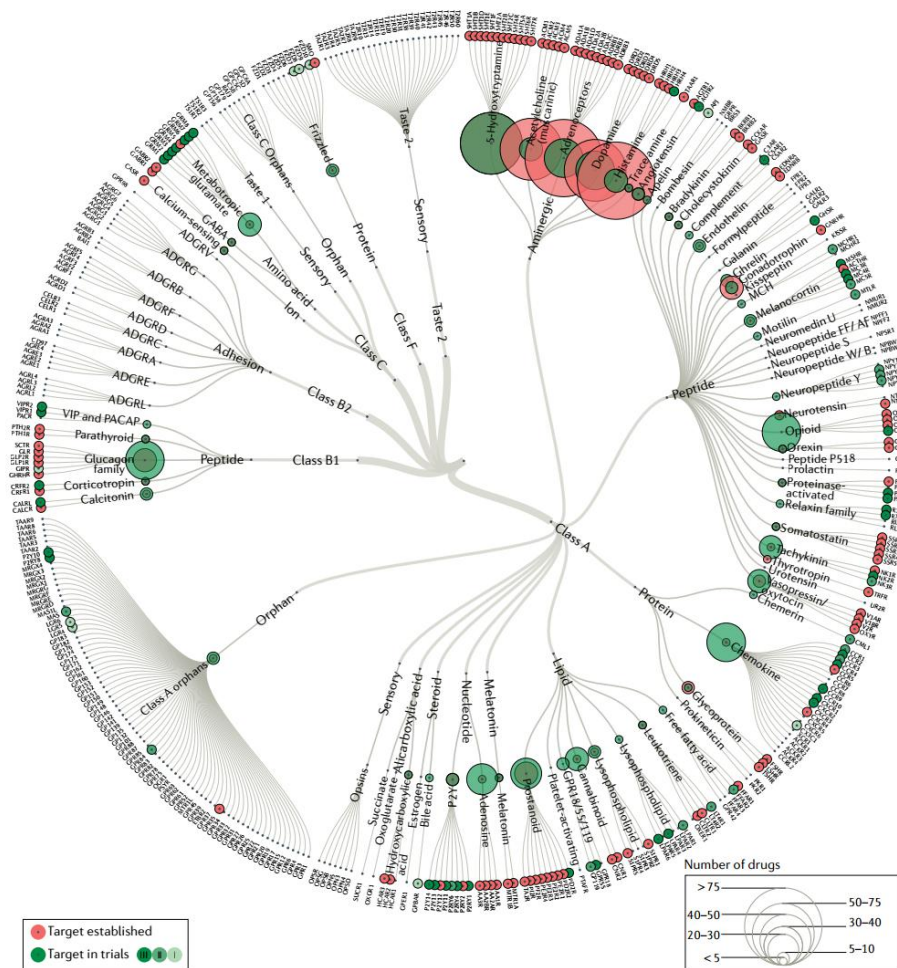
Inactive state

Antagonists

Rest state

Inhibitors

GPCR as Drug Targets



475 FDA approved drugs (~34%) act on 108 unique GPCR targets. 321 are in clinical trial.

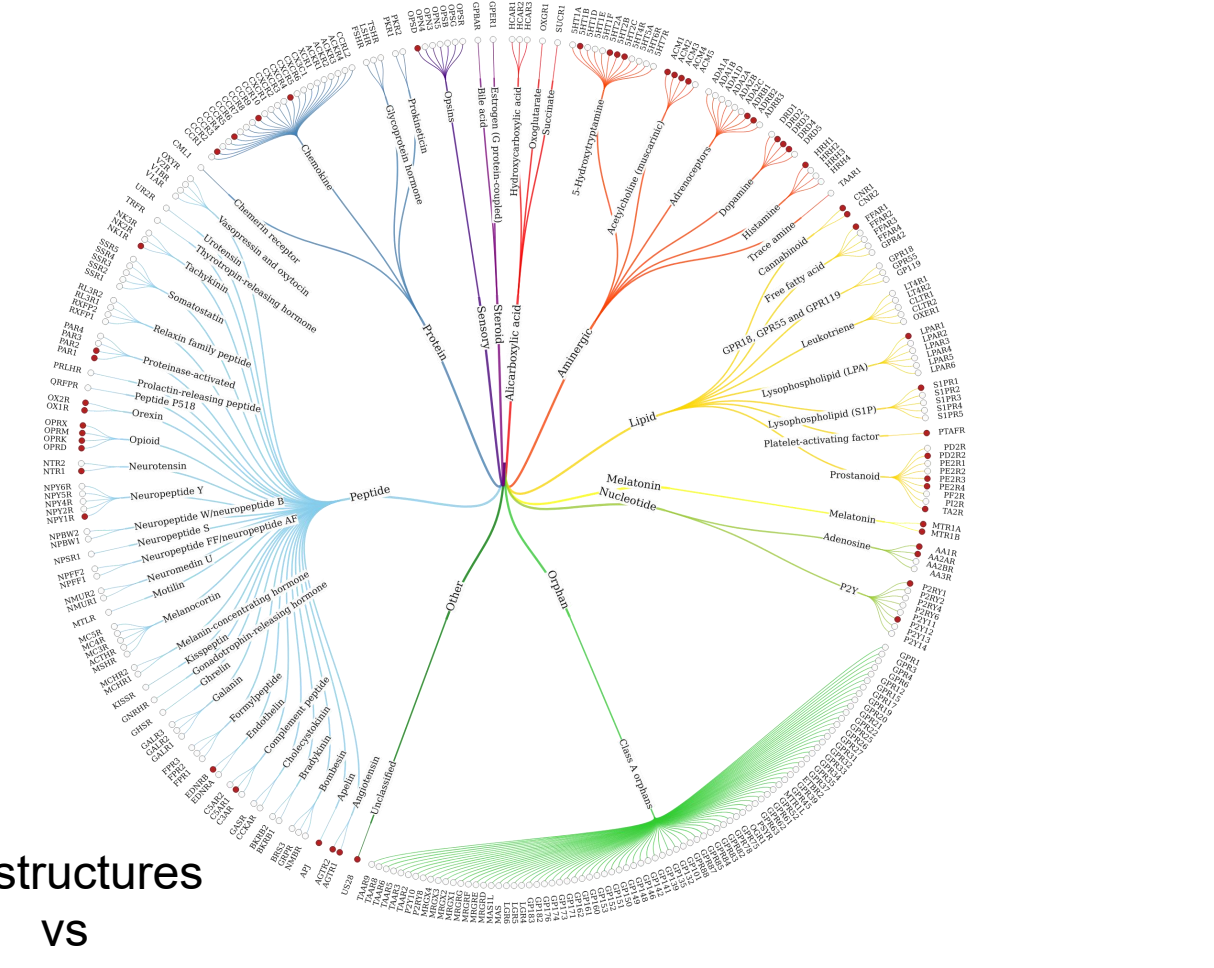
54 structures vs 800+ class A receptors

vs

800+ class A receptors

vs

A. S. House, et al. *Nature Review*, 2017, vol 16, 829 <https://gpcrdb.org/structure/statistics>



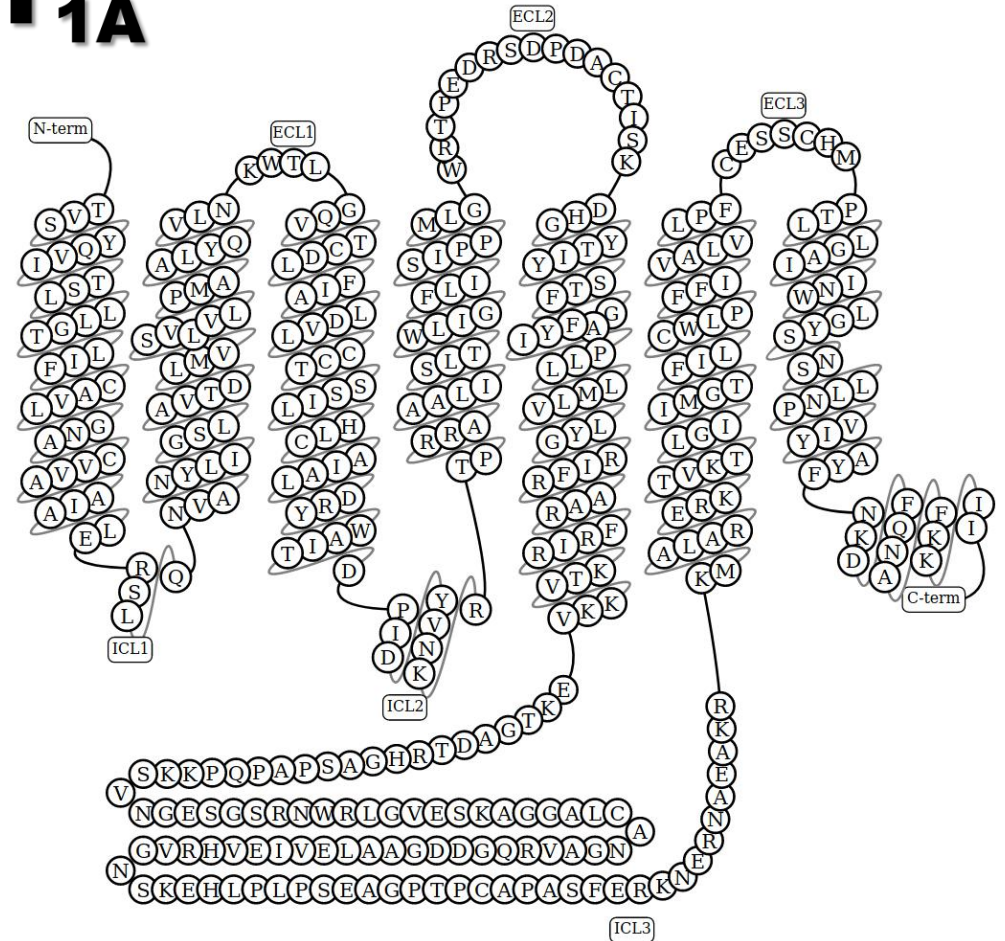
Serotonin Receptor 5-HT_{1A}

```
>sp|P08908|5HT1A_HUMAN 5-hydroxytryptamine receptor 1A
OS=Homo sapiens OX=9606 GN=HTR1A PE=1 SV=3
MDVLSPGQGNNNTTSPAPFETGGNTTGISDVTVSYQVITSLLLGLTIFCAVLGNACVVAA
IALERSLQNVANYLIGSLAVTDLMVSVLVLPMALYQVLNKWTLGQVTCDFIALDVLCC
TSSILHLCAIALDRYWAITDPIDYVNKRTPRRAAALISLTWLGFLISIPPMLGWRTPED
RSDPDACTISKDHGYTIYSTFGAFYIPLLLMLVLYGRIFRAARFRIRKTVKKVEKTGADT
RHGASPAPQP KKS VNGESGSRNWRLGVESKAGGALCANGAVRQDDGAALVIEVHRVGN
SKEHLPPLPSEAGPTPCAPASFERKNERNAEAKRKMALARERKT V K T L G I I M G T F I L C W L P
FFIVALVLPFCESCHMPTLLGAIINWLGYSNSLLNPVIYAYFNKDFQNAFKKI IKCKFC
RQ
```

Sequence

N-term			TM1		
MDVLSPGQGN	NTTSPAPFPE	TGGNTTGISD	VTVSYQVITS	LLLGLTIFCA	VLGNACVVAA
10	20	30	40	50	60
ICL1	TM2			ECL1 TM3	
IALERSLQNV	ANYLIGSLAV	TDLMVSVLVL	PMAALYQVLN	KWTLGQVTC	LFIALDVLCC
70	80	90	100	110	120
TSSILHLCAI	ALDRYWAITD	PIDYVNKRTP	RRAAALISLT	WLGFLISIP	PMLGWRTPED
130	140	150	160	170	180
RSDPDACTIS	KDHGYTIYST	FGAFYIPLLL	MLVLYGRIFR	AARFRIRKTV	KKVEKTGADT
190	200	210	220	230	240
RHGASPAPQP	KKS VNGESGS	RNWRLGVESK	AGGALCANGA	VRQDDGAAL	EVIEVHRVGN
250	260	270	280	290	300
SKEHLPPLPSE	AGPTPCAPAS	FERKNERNAE	AKRKMALARE	RKTVKTLGII	MGTFILCWLP
310	320	330	340	350	360
FFIVALVLPF	CESSCHMPTL	LGAIINWLG	SNSLLNPVIY	AYFNKDFQNA	FKKIICKFC
370	380	390	400	410	420
C-term					
R Q					

Domain Information



Diagram

But no 3D Structures yet!

Build 3D Structure of 5-HT_{1A} Receptor with Swiss-model

SWISS-MODEL

Modelling Repository Tools Documentation Log in Create Account

Welcome to SWISS-MODEL

SWISS-MODEL is a fully automated protein structure homology-modelling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

[Start Modelling](#)

2019 novel coronavirus (2019-nCoV), officially named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes COVID-19 (also known as 2019-nCoV acute respiratory disease). It is the cause of the ongoing 2019–20 Wuhan coronavirus outbreak, a global health emergency.

We modelled the full 2019-nCoV proteome based on the NCBI reference sequence [NC_045512](#) and annotations from [UniProt](#).

The results are available [here](#).

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in [SWISS-MODEL Repository](#)?

Search SWISS-MODEL Repository

<https://swissmodel.expasy.org/>

SWISS-MODEL Interactive Workspace - Mozilla Firefox

SWISS-MODEL

Modelling Repository Tools Documentation Log in Create Account

Start a New Modelling Project

Target Sequence(s):
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

Paste your target sequence(s) or UniProtKB AC here

[+ Upload Target Sequence File...](#) [Validate](#)

Project Title:

Email:

[Search For Templates](#) [Build Model](#)

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

You are currently not logged in - to take advantage of the workspace, please [log in](#) or [create an account](#).

(There is no requirement to create an account to use any part of SWISS-MODEL, however you will gain the benefit of seeing a list of your previous modelling projects here.)

Modelling Projects in Session

- [Untitled Project](#)

Created: today

Build 3D Structure of 5HT_{1A} Receptor with Swiss-model

Untitled Project | Templates - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Untitled Project | Temp x

https://swissmodel.expasy.org/inter

BIOZENTRUM University of Basel The Center for Molecular Life Sciences

SWISS-MODEL Modelling Repository Tools Documentation Log in Create Account

All Projects

Untitled Project Created: today at 18:32

Summary Templates 50 Models

Template Results


Templates Quaternary Structure Sequence Similarity

Alignment of Selected Templates More

Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method
<input checked="" type="checkbox"/>	5v54.1.A	5-hydroxytryptamine receptor 1B,OB-1 fused 5-HT1b receptor,5-hydroxytryptamine receptor 1B		0.70	0.32	42.97	X-ray, 3.9Å
<input type="checkbox"/>	5v54.1.B	5-hydroxytryptamine receptor 1B,OB-1 fused 5-HT1b receptor,5-hydroxytryptamine receptor 1B		0.70	0.32	42.97	X-ray, 3.9Å
<input type="checkbox"/>	4iar.1.A	Chimera protein of human 5-hydroxytryptamine receptor 1B and E. Coli soluble cytochrome b562		0.68	-	41.60	X-ray, 2.7Å
<input type="checkbox"/>	4iaq.1.A	Chimera protein of human 5-hydroxytryptamine receptor 1B and E. Coli soluble cytochrome		0.67	0.23	41.22	X-ray, 2.8Å

Build Models 1

Clear Selection



BIOZENTRUM University of Basel The Center for Molecular Life Sciences

SWISS-MODEL Modelling Repository Tools Documentation Log in Create Account

All Projects

Untitled Project Created: today at 18:32

Summary Templates 50 Models 1

Model Results Order by: GMQE

Oligo-State: Monomer (matching prediction)

Ligands: None

GMQE: 0.68 QMEAN: -4.48

Global Quality Estimate: QMEAN -4.48, Cβ -4.21, All Atom -1.29, solvation 0.75, torsion -4.19

Local Quality Estimate:

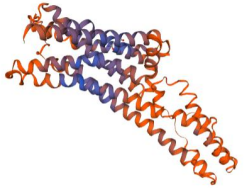
Comparison:

Template	Seq Identity	Coverage	Description
5v54.1.A	42.97%		5-hydroxytryptamine receptor 1B,OB-1 fused 5-HT1b receptor,5-hydroxytryptamine receptor 1B

Model-Template Alignment

```

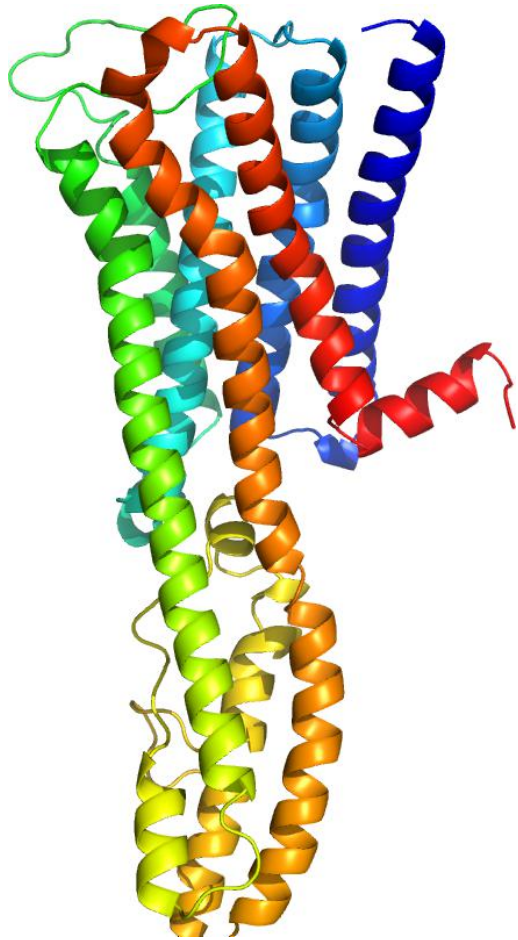
Model_01 MDVLSPPGQGNNTSPPAPPFTGGNTTGISDVTIVSYQVITSLLLGTLIFCAVLGNA 55
5v54.1.A -----IS(CFWKLVLLVPLLLALITLATTLSNA 32
Model_01 CVVAATALEERSLQNVANYLIGSLAVTDLMVSYLVLPMAALYQVLNKNWTLGQVTCG 110
5v54.1.A FVIATVYRKRKLRHTQANYLIRSLAVTDLLVSLVPISTHYIVYGRWTLGQVTCG 87
Model_01 LFIALDVLCCCTSSILHLCAIALDRYWAITDPIDYVNKRTPRRAAALISLTLWLGIF 165
5v54.1.A FVLSDDITCCCTASIHVLCVIALDRYWAITDAVEYSARKRTPRAAVIALVWVFSI 142
Model_01 LISPPMLGWRTPEDRSDPDACTISKDHGYTIYSTFGAFYIPLLLMLVLYGRIF 219
5v54.1.A EISLPPFFWRQAKAEVSECVVNTDHLTYVYSTVGAFFPTLLLLIYALYGRIV 196
Model_01 RAARFRIRKTVKVEKTGADTRHGASPAQPQKKSVM---GESGRNWRGLGV 267
5v54.1.A VEARSRIADLEDRWRTELDNLRVJEDAAANAQAEVREALTRRRAAEDARATPAPAL 251
Model_01 ESKAGGALCANGAVRQDDGAALEVIEVHRVGNKEHLPLPSEAGPTPCAPASFE 322
5v54.1.A DRSPASPEMEDFRHGFDTL-----VGQIDDALRLADBSRVAFAQAQAAEE 296
Model_01 RKNERNAEAKRKNALARERKTKVKTGLQIMGTFLICLWLPFFIVALVLPFCESCHW 377
5v54.1.A LRTTRNAYIQYLNARERKATKTLGIILGAFIWCWLPFFIISLVPICKDACWF 351
Model_01 PTLGAIINWLGYSNSLLNPVIYAYFNKDFQNAFKKIKKFCRQ 422
5v54.1.A HLAIFDFFTLWGLVNSLNPVIYTHNDFKQAFHKIIRFK----- 392
                    
```



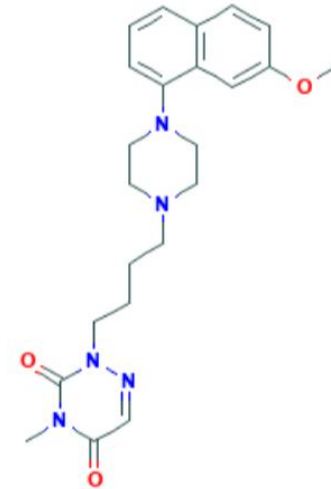
NGL Cartoon

1 422

Molecular Docking with Autodock on BioHPC



Homology model of 5-HT_{1A} Receptor



PubChem ID: 10324985
EC50 with 5-HT_{1A}: 0.47 nM

Molecular Docking with Autodock on BioHPC

<https://portal.biohpc.swmed.edu/>

-> Cloud Services

-> Web Visualization

-> VNC Viewer

The image shows a composite of two screenshots. The top screenshot is a Mozilla Firefox browser window displaying the BioHPC portal website. The website header includes 'UT Southwestern Medical Center BioHPC' and 'Lydia Hill Department of Bioinformatics'. The main content area is titled 'Web based Visualization' and lists various services like WebGUI, WebGPU, WebGPU4, WebWinDCV, and WebDIGITS. A terminal window is visible at the bottom of the browser page with the URL <https://portal.biohpc.swmed.edu/terminal/ssh/>.

The bottom screenshot is a VNC viewer window titled 'x11 [Tight + JPEG 1X Q95 + CL 1]'. It displays the AutoDockTools (ADT) interface. The interface includes a menu bar (File, Edit, View, Search, Terminal, Help), a toolbar, and a main 3D graphics window showing a protein structure with a ligand docked. A terminal window is overlaid on the ADT interface, showing the following text:

```
File Edit View Search Terminal Help
/programs/share/LICENSE. The app
exclusively to member laboratorie
*****
SBGrid was developed with support
HHMI, and NSF. If use of SBGrid c
in your publication, please inclu
*****
Software used in the project was
cite: eLife 2013;2:e01456. Collab
*****
SBGrid installation last updated:
Please submit bug reports and hel
*****
Capsule Status: Active
*****
For additional information
*****
[s190450@NucleusA203 ~]$ adt
setting PYTHONHOME environment
Run ADT from /programs/x86_64-Lin
MSMSLIB 1.4.4 started on NucleusA2
Copyright M.F. Sammer (March 2006)
Compilation flags
```


Molecular Docking with Autodock on BioHPC

Add sbgrid to your environment

vi .bashrc

. /programs/sbgrid.shrc

source .bashrc

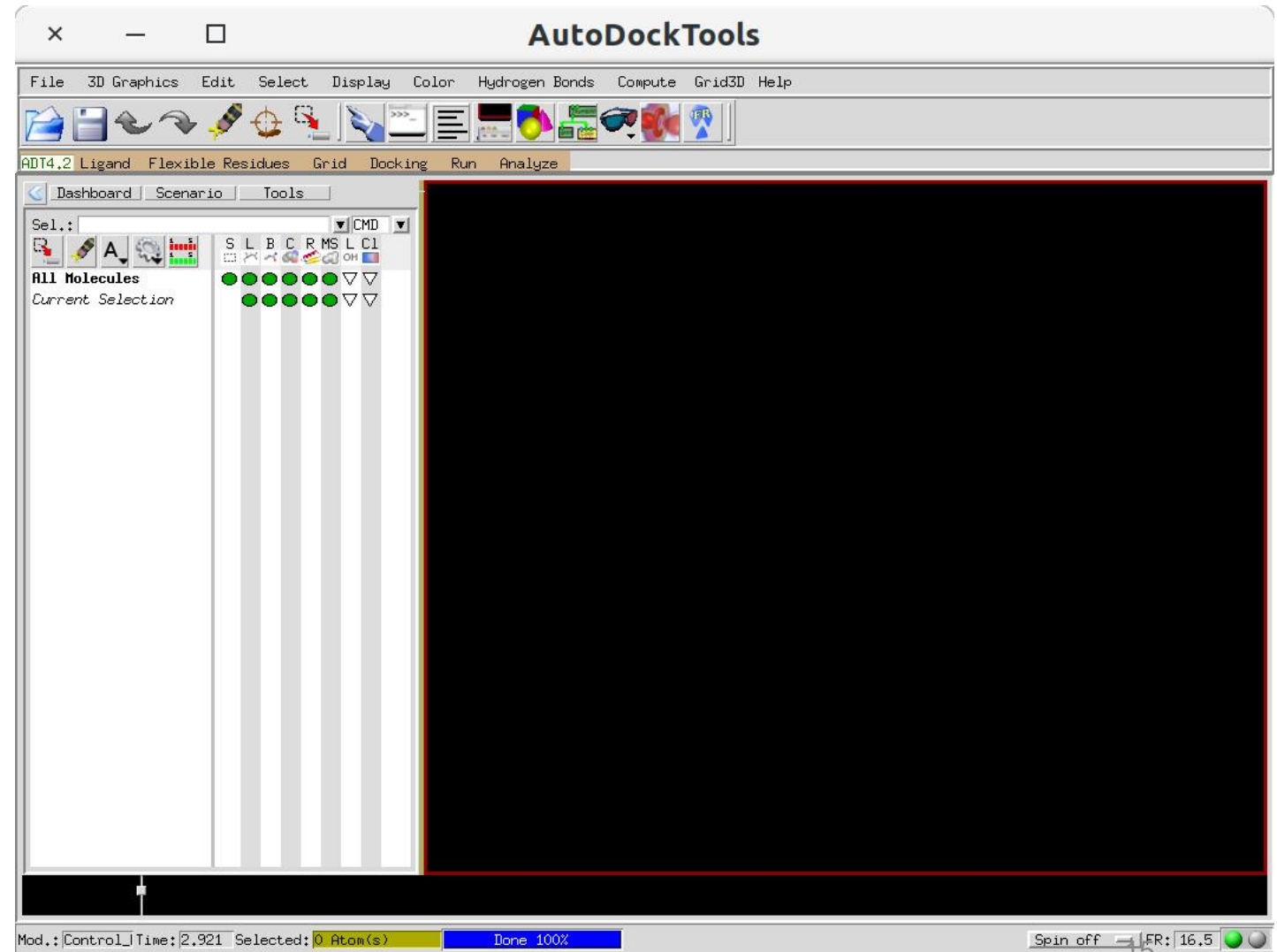
```
[s190450@Nucleus005 ~]$ vi .bashrc
[s190450@Nucleus005 ~]$ source .bashrc
*****
                          Software Support by SBGrid (www.sbgrid.org)
*****
                          SBGrid Announcements
- SBGrid is not compatible with MacOS 10.15 Catalina.
  see https://sbgrid.org/wiki/catalina for more info.
*****
Your use of the applications contained in the /programs directory constitutes
acceptance of the terms of the SBGrid License Agreement included in the file
/programs/share/LICENSE. The applications distributed by SBGrid are licensed
exclusively to member laboratories of the SBGrid Consortium.
*****
SBGrid was developed with support from its members, Harvard Medical School,
HHMI, and NSF. If use of SBGrid compiled software was an important element
in your publication, please include the following reference in your work:

Software used in the project was installed and configured by SBGrid.
cite: eLife 2013;2:e01456, Collaboration gets the most out of software.
*****
SBGrid installation last updated: in-progress (Update available)
Please submit bug reports and help requests to:      <bugs@sbgrid.org> or
                                                    <http://sbgrid.org/bugs>
*****
Capsule Status: Active
For additional information visit https://sbgrid.org/wiki/capsules
*****
[s190450@Nucleus005 ~]$ █
```

Molecular Docking with Autodock on BioHPC

```
# start Autodock Tools
```

```
adt
```



Molecular Docking with Autodock on BioHPC

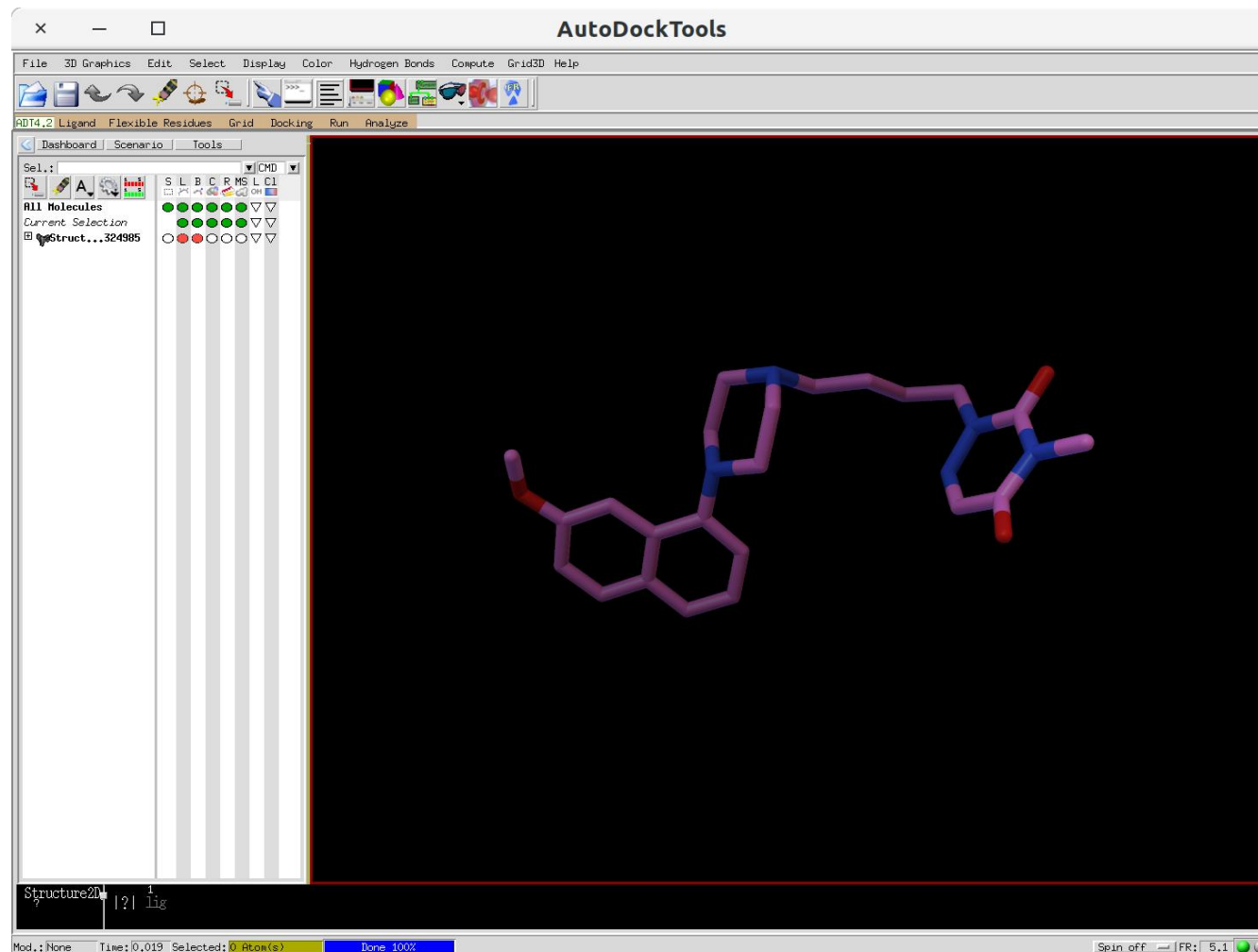
Prepare the **ligand**

Ligand -> Input -> Open -> lig.pdb

Ligand -> Torsion Tree -> Choose Torsions

Ligand -> Output -> Save as **lig.pdbqt**

Edit -> Delete -> Delete all Molecules



Molecular Docking with Autodock on BioHPC

Prepare the **receptor**

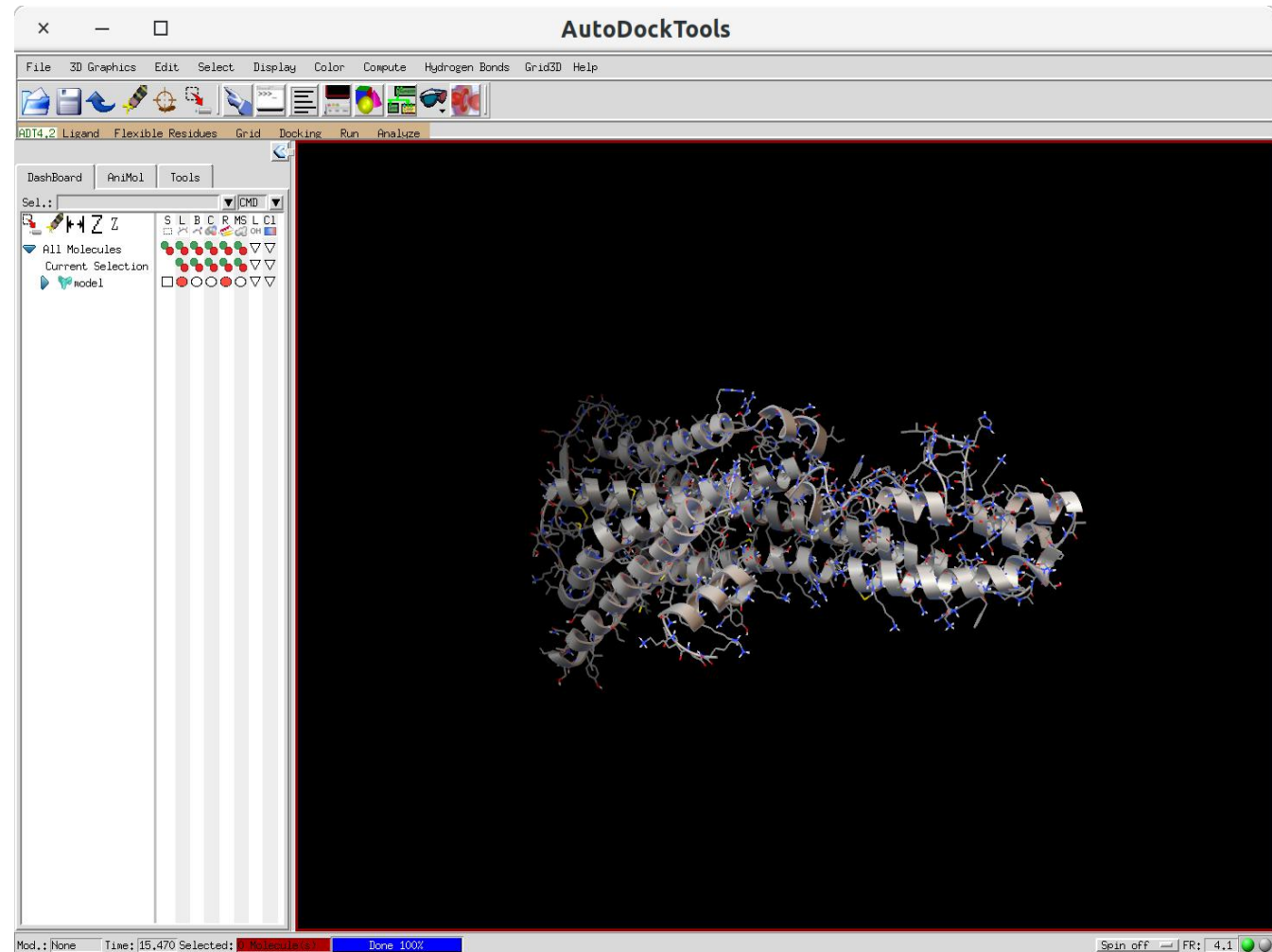
File-> Read Molecule -> model.pdb

Edit-> Hydrogens -> Add -> Polar only

Grid-> Macromolecule -> Choose ->

Save as **model.pdbqt**

Edit -> Delete -> Delete all Molecules



Molecular Docking with Autodock on BioHPC

Prepare the **grids**

Grid -> Macromolecule -> Open -> **model.pdbqt**

Grid -> Set Map Types -> Open ligand -> **lig.pdbqt**

Grid -> Grid Box -> (File -> Close saving current)

Grid -> Output -> **box.gpf**

run **autogrid4**

autogrid4 -p box.gpf -l autogrid.log

(Run -> Run AutoGrid)

output files

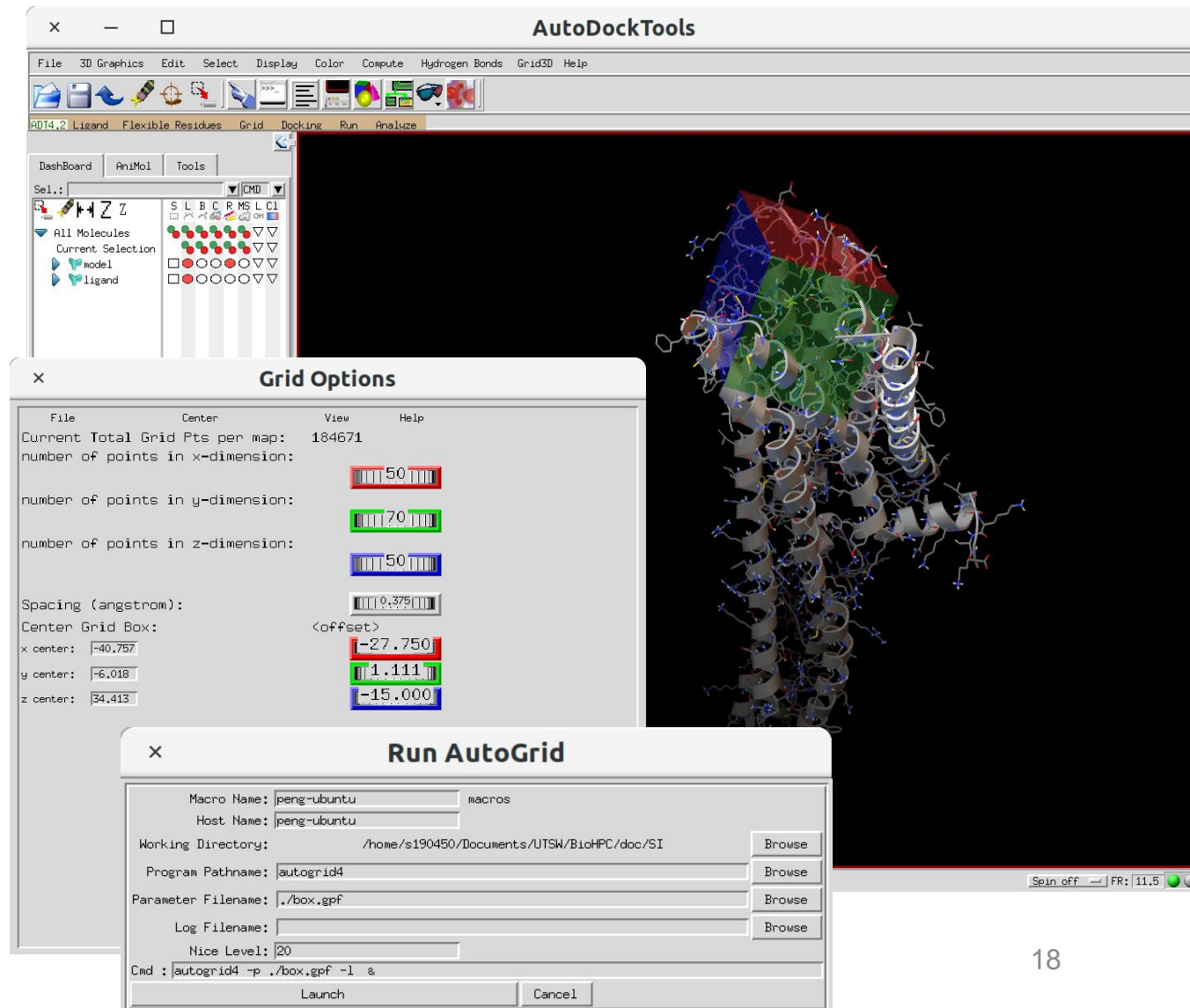
model.A.map model.d.map

model.maps.fld model.NA.map

model.OA.map model.C.map

model.e.map model.maps.xyz

model.N.map



Molecular Docking with Autodock on BioHPC

Docking

Docking -> Macromolecule ->

Set Rigid Filename, **rigid.pdbqt**

Docking -> Ligand -> Open -> **lig.pdbqt**

Docking -> Search Parameters -> Genetic Algorithm

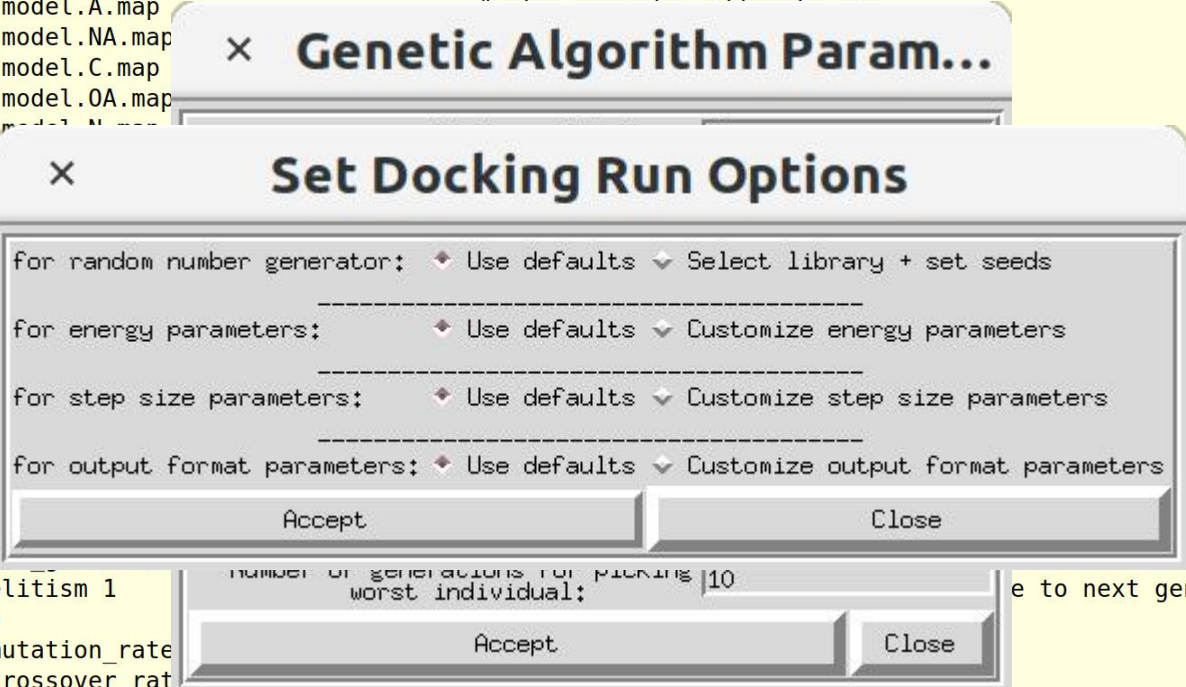
Docking -> Docking Parameters

Docking -> Output -> Lamarckian GA(4.2) -> **dock.dpf**

autodock4 -p dock.dpf -l dock.dlg

(Run -> Run AutoDock)

```
autodock_parameter_version 4.2      # used by autodock to validate parameter set
outlev 1                            # diagnostic output level
intelec                             # calculate internal electrostatics
seed pid time                        # seeds for random generator
ligand_types A NA C OA N           # atoms types in ligand
fld model.maps.fld                  # grid_data_file
map model.A.map                      #
map model.NA.map                     #
map model.C.map                       #
map model.OA.map                      #
elec                                #
desc                                #
move                                #
about                                #
tran                                #
quat                                #
dihedr                              #
torsions                             #
rmsd                                 #
extra                                #
e0max                                #
ga_elitism 1                         #
ga_mutation_rate                    #
ga_crossover_rate                   #
ga_window_size 10                   #
ga_cauchy_alpha 0.0                 # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0                  # Beta parameter Cauchy distribution
set_ga                               # set the above parameters for GA or LGA
sw max its 300                       # iterations of Solis & Wets local search
```



Molecular Docking with Autodock on BioHPC

Analysis

Analyze -> Docking -> Open -> **dock.dlg**

Analyze -> Conformations -> Load -> **dock.dlg**

autodock Conformation Ch...

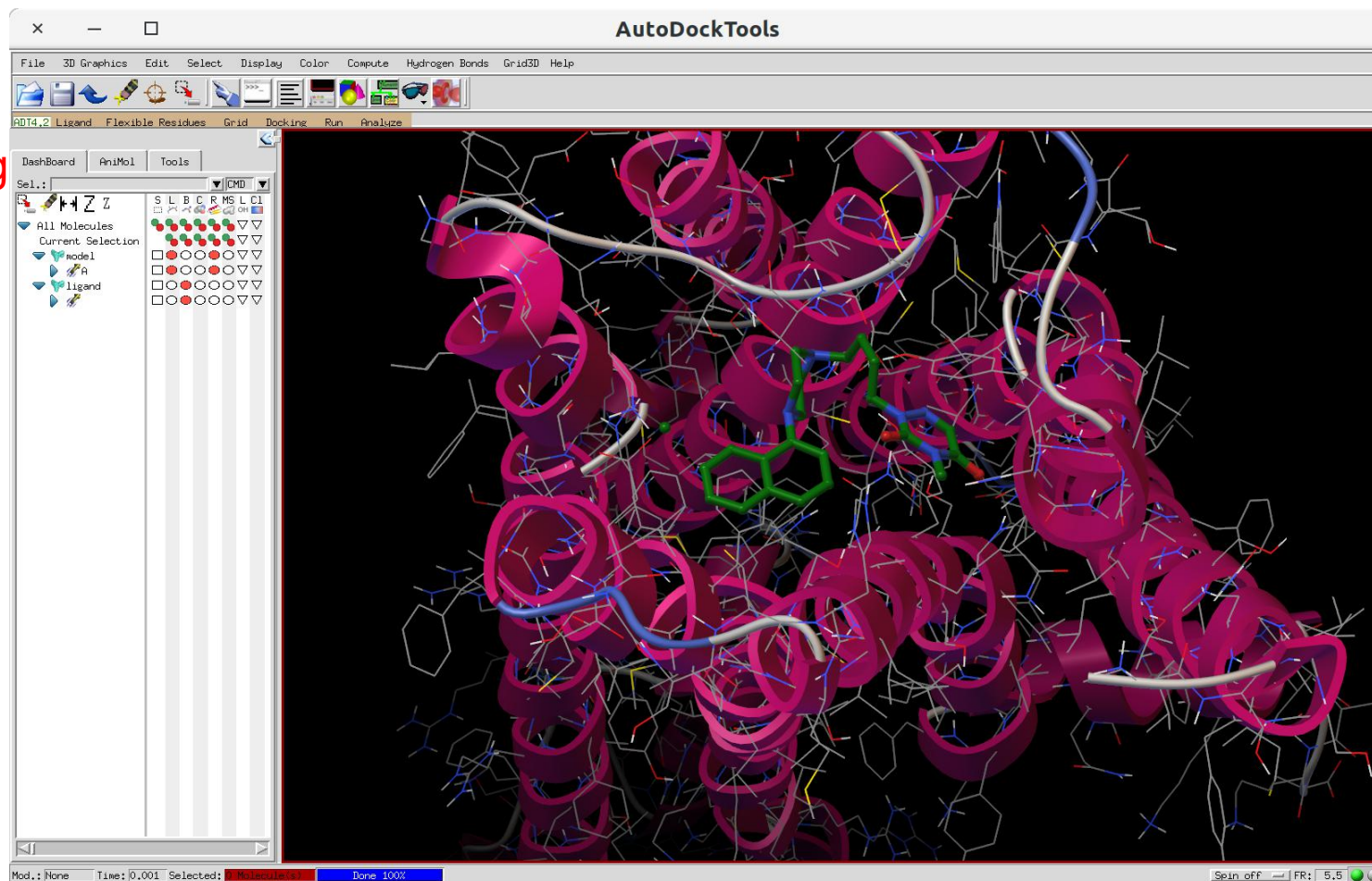
Rank: 1_1
Binding Energy: -8.14
KI : 1.09uM
Intermolecular Energy : -10.22
Internal Energy : -1.53
Torsional Energy : 2.09
Unbound Extended Energy: -1.53
Cluster RMS: 0.0
Ref RMS: 52.12

select from 10 dockings:
(double click to update coords)
(Rank_SubRank docked energy)

ligand input	
ligand 1_1	-8.14
ligand 1_2	-7.48
ligand 1_3	-7.11
ligand 2_1	-7.77
ligand 2_2	-7.09
ligand 2_3	-6.65
ligand 2_4	-6.19
ligand 3_1	-7.29
ligand 4_1	-6.83

Write Current Coords

Dismiss



Run on Your PC

AutoDockTools

http://autodock.scripps.edu/downloads/resources/adt/index_html

AutoDock

<http://autodock.scripps.edu/downloads/autodock-registration/autodock-4-2-download-page/>

Virtual Screening



ZINC15

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains **over 750 million** purchasable compounds you can search for analogs in under a minute.

ZINC is provided by the [Irwin](#) and [Shoichet](#) Laboratories in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). We thank [NIGMS](#) for financial support (GM71896).

To cite ZINC, please reference: Sterling and Irwin, *J. Chem. Inf. Model*, 2015 <http://pubs.acs.org/doi/abs/10.1021/acs.jcim.5b00559>. You may also wish to cite our previous papers: Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model*, 2012 DOI: [10.1021/ci3001277](https://doi.org/10.1021/ci3001277) or Irwin and Shoichet, *J. Chem. Inf. Model*. 2005;45(1):177-82 [PDF](#), [DOI](#).

Getting Started

- [Getting Started](#)
- [What's New](#)
- [About ZINC 15 Resources](#)
- [Current Status / In Progress](#)
- [Why are ZINC results "estimates"?](#)

Explore Resources

Ask Questions

You can use ZINC for **general** questions such as

- [How many substances in current clinical trials have PAINS patterns? \(150\)](#)
- [How many natural products have names in ZINC and are not for sale? \(9296\) get them as SMILES, names and calculated](#)

ZINC15 News

- 2018-02-14 - ZINC reaches 213,235,528 purchasable leadlike 3D!
- 2018-02-13 - ZINC reaches 736,001,654 purchasable molecules 2D!
- 2018-01-14 - Klara Anu is born! Welcome Klara Anu, sister to Lisa!
- 2018-01-01 - Chinzo Dandar joins our team. Welcome Chinzo! Follow us on

Thanks for your attention!

biohpc-help@utsouthwestern.edu